A Novel Approach to Spatially Indexed Functional Data Analysis

Luke A Barratt, John AD Aston

Statistical Laboratory, University of Cambridge, Cambridge, United Kingdom

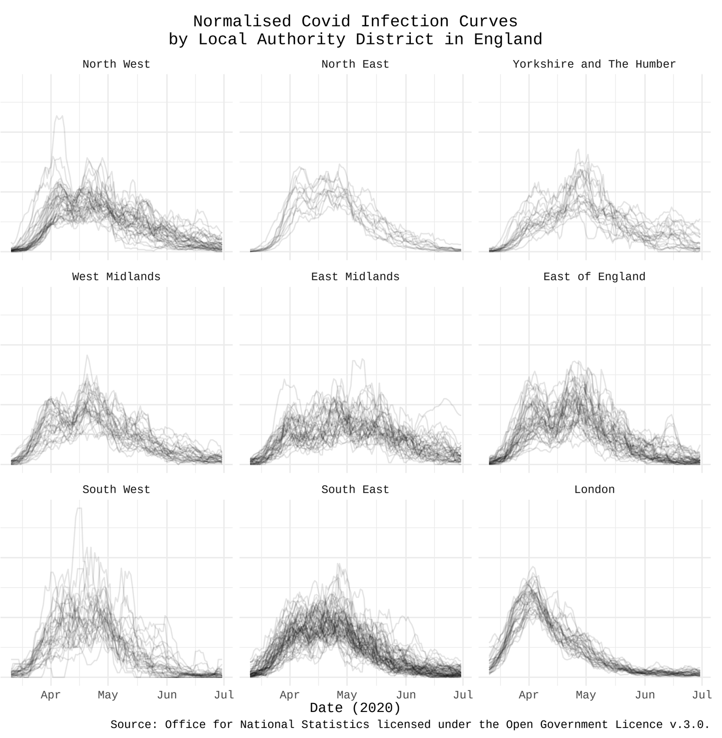
The Problem

We suppose we are observing functional data (see [1]) indexed by spatial locations, and we wish to understand the phase variation of the curves across space. In particular, we consider the problem of the registration or alignment of spatially correlated functional data in order to understand the warping away from a theoretical typical trajectory. (See [2] for the iid case.)

We in particular focus on two data sets:

- daily Covid infection data at the LAD level in England; and
- weekly death data at the NUTS 3 level in Europe.

We use these data as a proxy for the waves of the Covid-19 pandemic passing through Europe.



The spatial observation of functional data is also common in many other fields, including meteorology, economics, neuroscience and ecology. We envisage myriad applications of a spatially aware methodology for functional registration, noting there is little prior work on this (although, see [3]).

The Model

We build our model of our functional data from two stationary random functional fields, in the same vein as [4], taking into account [5]:

- The variation in amplitude is understood through the random functional field X, representing the overall shape of the observable curve, its values $X_u : [0,1] \longrightarrow \mathbb{R}$. We assume it to have predominantly rank-one variation:
$$X_u(t) = \xi_u \mu(t) + \delta \epsilon_u(t), \tag{1}$$
for constant and unit-norm μ, random stationary scalar field ξ, small constant $\delta << \sqrt{\mathrm{Var}\,\xi_u}$, and random unit-norm functional field ϵ.
- The variation in phase is understood through the random functional field h, which represents the warping from said overall shape the observable curve exhibits, its values $h_u : [0,1] \longrightarrow [0,1]$. It is only assumed that $h_u(0) = 0$, h_u is a diffeomorphism, and $\mathbb{E}h_u = \mathrm{id}$.

Putting these together, we observe (possibly with measurement error) from the random functional field Y:
$$Y_u(t) = X_u(h_u^{-1}(t)). \tag{2}$$

The Method

We observe Y (possibly with measurement error) at spatial locations $(u_i)_{i=1}^n$ and temporal locations $(t_j)_{j=1}^m$. From these data we estimate the h_i thus:

1. For each i, we extend our discrete observations to continuous curves by regression, and normalise these to make them comparable.
2. For each pair (i,k) $(i \neq k)$, we then estimate the pairwise warping $g_{ik} := h_i \circ h_k^{-1}$ by minimising the following target:
$$\|Y_i \circ g - Y_k\|^2 + \lambda \|g - \mathrm{id}\|^2, \tag{3}$$
over all g in some finite-dimensional approximation of all valid warping functions, and for some regularisation constant λ.
3. For each i, we estimate a functional variogram (cf. [6]),
$$2\gamma_i(u_k, u_l) = \mathbb{E}\|g_{ki} - g_{li}\|^2, \tag{4}$$
by making the approximation $2\gamma_i(u_k,u_l) \approx 2\widetilde{\gamma}_i(d(u_k,u_l))$ for some function $\widetilde{\gamma}_i$; we do this by a standard parametric regression for variogram estimation.
4. We then utilise the approximation
$$\mathbb{E}_i \langle g_{ki} - \mathbb{E}_i g_{ki}, g_{li} - \mathbb{E}_i g_{li} \rangle \approx \widetilde{\gamma}_i(\infty) - \widetilde{\gamma}_i(d(u_k,u_l)) =: \widehat{C}_{kl}^{(i)}, \tag{5}$$
where $\mathbb{E}_i$ is shorthand for taking an expectation conditional on h_i, in order to estimate approximate weights:
$$\widehat{w}^{(i)} \propto (\widehat{C}^{(i)})^{-1} 1_n, \tag{6}$$
where 1_n is a length-n vector of ones, and $\sum_k \widehat{w}_k^{(i)} = 1$.
5. We then finally estimate the h_i as a weighted mean of the $(g_{ki})_k$:
$$\widehat{h}_i^{-1} = \sum_{k \neq i} \widehat{w}_k^{(i)} g_{ki}. \tag{7}$$

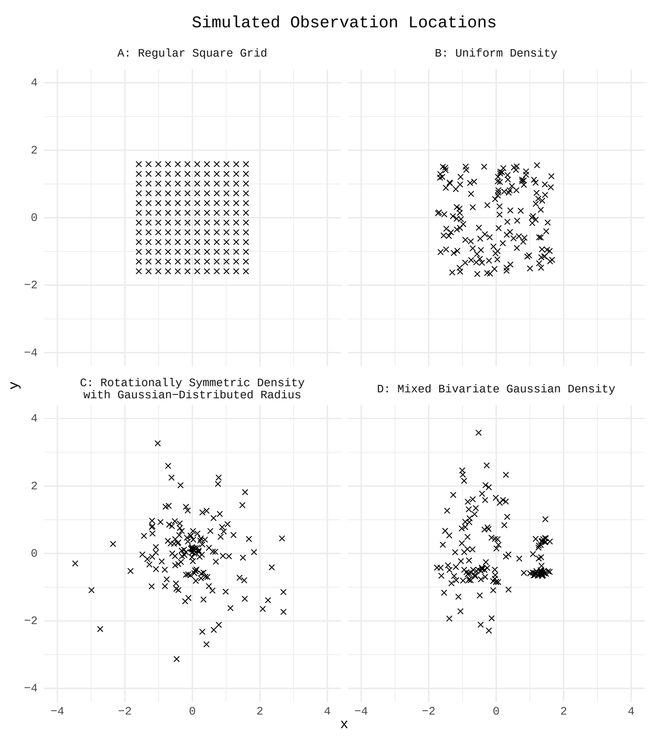
These weights are approximately optimal in the sense of minimising the MSE $\mathbb{E}\|\widehat{h}_i - h_i\|^2$.

This follows standard approaches to spatial data analysis (see [7]).

Simulations

Simulations were run to estimate the MSE in the estimation of the h_i, under a variety of observation location sets (see below), covariance structures for warping functions and amplitude models. We define the MSE as:
$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\|\widehat{h}_i^{-1} - h_i^{-1}\|^2 \tag{8}$$

It was found that the MSE was almost always reduced when compared to a simple average methodology (a non-spatial approach; see [4]), indeed up to by a factor of two. The results are also given below.



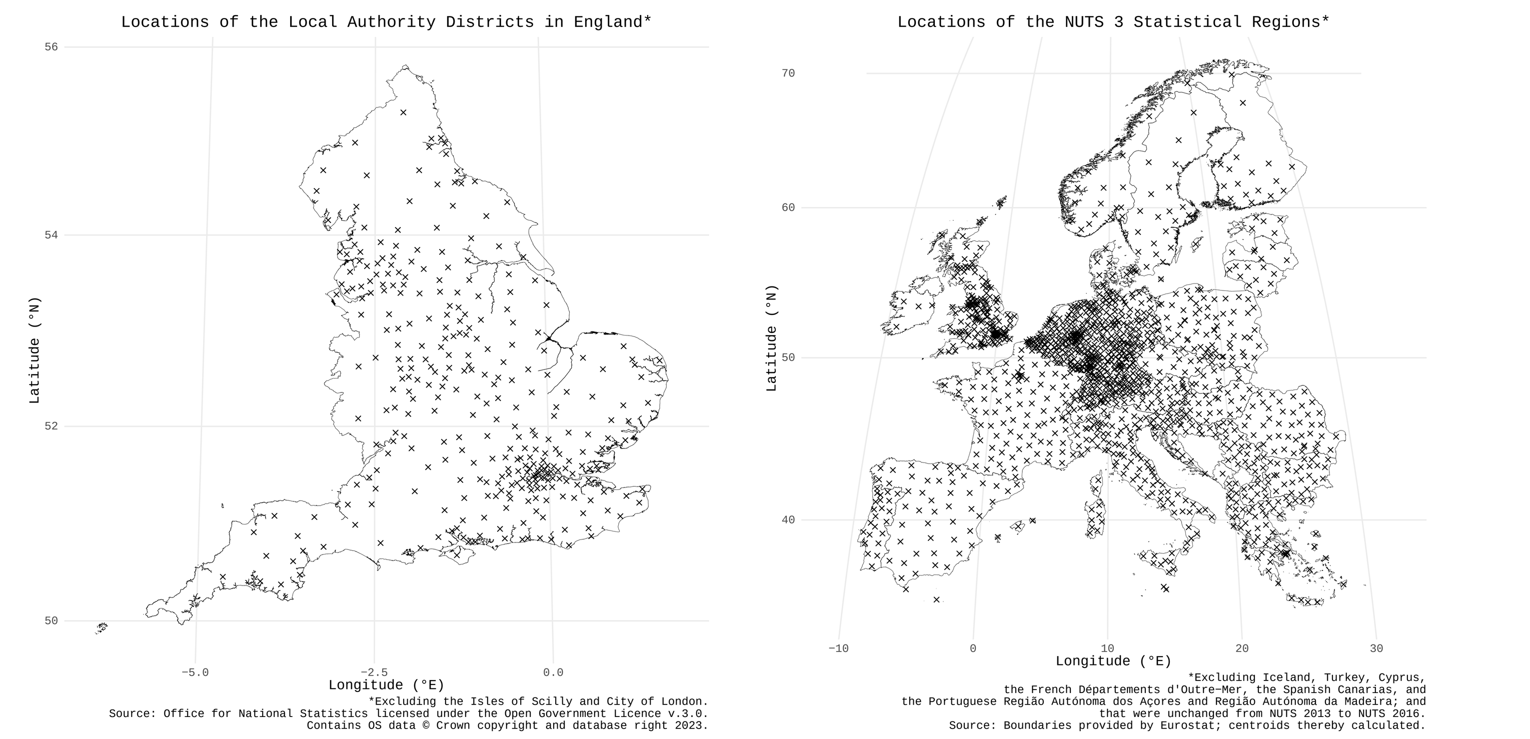
Simulation Results

Locations	Simple	Weighted	Impr.
A	0.00102	0.00086	16.0%
B	0.00103	0.00089	13.4%
C	0.00109	0.00069	36.9%
D	0.00106	0.00082	22.6%

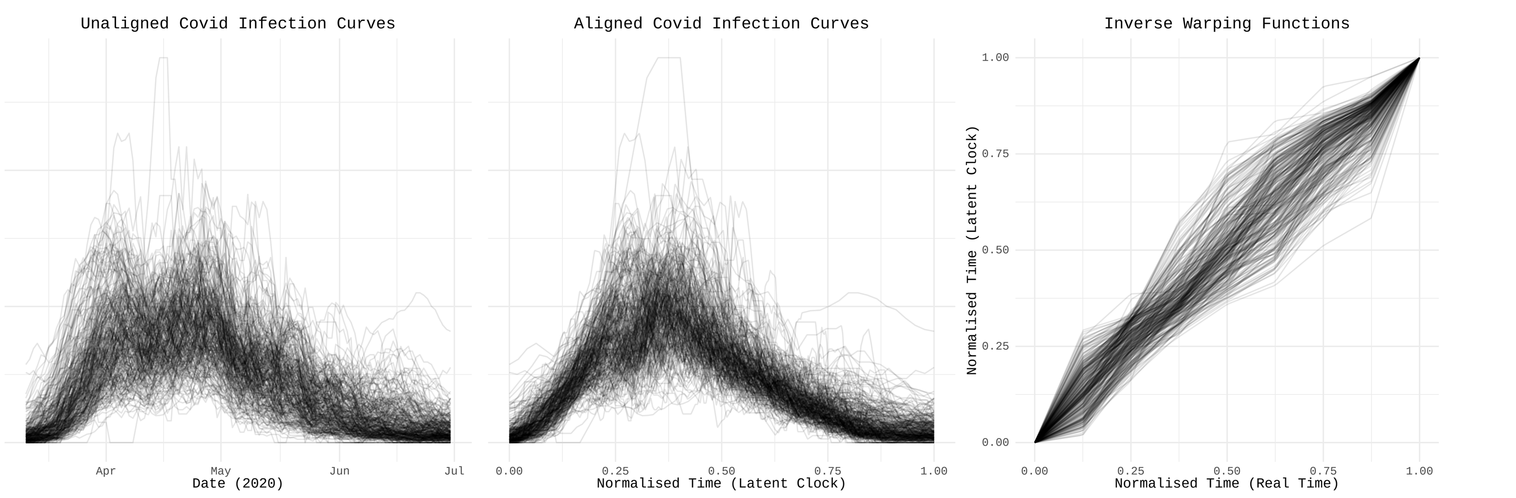
The data provided are estimates of mean squared error under the traditional, non-spatial (simple mean) and our spatial (weighted mean) methodologies. The observation locations correspond to the figure on the left.

Applications

The developed methodology was applied to the first Covid wave of 2020 in English LADs and European NUTS 3 regions. The distribution of these regions are illustrated below.



We in particular consider here the application to English LADs. The resulting aligned curves and warping functions are given below. With the aligned curves and warping functions, we can now perform separate spatially-aware functional principal component analysis, for example, to understand the variation further.



References

[1] J. Ramsay and B. W. Silverman, *Functional data analysis* (Springer Series in Statistics). Springer, 1997.

[2] A. Srivastava, W. Wu, S. Kurtek, E. Klassen, and J. S. Marron, “Registration of functional data using fisher-rao metric,”

[3] X. Guo, S. Kurtek, and K. Bharath, “Variograms for kriging and clustering of spatial functional data with phase variation,” *Spatial statistics*, vol. 51, p. 100 687, 2022.

[4] R. Tang and H.-G. Müller, “Pairwise curve synchronization for functional data,” *Biometrika*, vol. 95, no. 4, pp. 875–889, 2008.

[5] A. Chakraborty and V. M. Panaretos, “Functional registration and local variations: Identifiability, rank, and tuning,” 2021.

[6] O. Gromenko, P. Kokoszka, L. Zhu, and J. Sojka, “Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends,” *The Annals of Applied Statistics*, pp. 669–696, 2012.

[7] N. Cressie, *Statistics for spatial data*. John Wiley & Sons, 1993.

Luke A Barratt: lab85@cam.ac.uk John AD Aston: jada2@cam.ac.uk